

# CA Water Board Data Management Strategy and Open Data Initiative



Greg Gearheart, Deputy Director  
Office of Information Management and Analysis

For WQCC - October 26, 2017

# Open Data Culture



# Office of Information Management and Analysis (OIMA)

## SWAMP Unit

**Lori Webber** (Sr ES)

Surface Water Ambient  
Monitoring Program -  
Overall Program  
Management and Support

## SWAMP IQ Unit

**Melissa Morris** (Sr ES)

Quality Assurance and  
Data Management  
Support for SWAMP and  
the CA Env Data  
Exchange Network  
(CEDEN)

## Data Integration Unit

**Jarma Bennett** (Sr Eng)

Data integration and  
analysis, CEDEN and  
CIWQS Project  
Management, “Open  
Data” Initiative

# Additional OIMA Functions and Staff (cont.)

- Citizen Monitoring, Open Science and SWAMP Southern CA Support - Erickson Burres (ES Spec)
- Chief Data Scientist, AB1755 Lead and Water Economist - Rafael Maestu (RPS2)
- Performance Management and Watershed Outcomes - Bev Anderson-Abbs (ES Spec)
- Water Quality Monitoring Council Director - vacant (EPM I)
- Quality Officer, Quality Management Program Manager - Renee Spears (ES Spec)
- SWAMP Assessment and Reporting Lead Scientist (pesticides, CECs and others) - Dawit Tadesse (ES Spec)

# Data Management Strategy (2017 update)

## Annual Civic Engagement Events

- Data Fair (open house)
- Data Innovation Challenge (hackathons)
- Water / Data Science Synthesis
- Brown Bag Series of Speakers
- Other partnerships

## Our Data Management Strategy Framework

- Based on Principles
- Will guide:
  - Data driven management
  - “Water” decisions
  - “Technology” decisions
  - Quality program
- Lists our data management values
- Encourages “data literacy”

# Databases and Datasets at the CA Water Boards

- Over 20 enterprise database applications
- Water quality, water rights, drinking water, etc.
- Program data (e.g, facilities, activities) and environmental / ambient data (e.g., surface water and groundwater quality, water use, water conservation, etc.)
- 18 data resources on [data.ca.gov](https://data.ca.gov) → more all the time

# Why do we collect, use and produce data?

- To inform our data-driven management and planning activities – performance report cards, workplans, resource assignment/augmentation, evaluating program effectiveness, and many others examples;
- To inform our critical decisions regarding our mission(s) and water management responsibilities – water allocation and use, water quality planning and “policies,” permitting, program prioritization, and many other examples; and
- To provide transparency to our many partners and stakeholders for their use, interests and purposes.

# Five Guiding Principles for WB Data Management

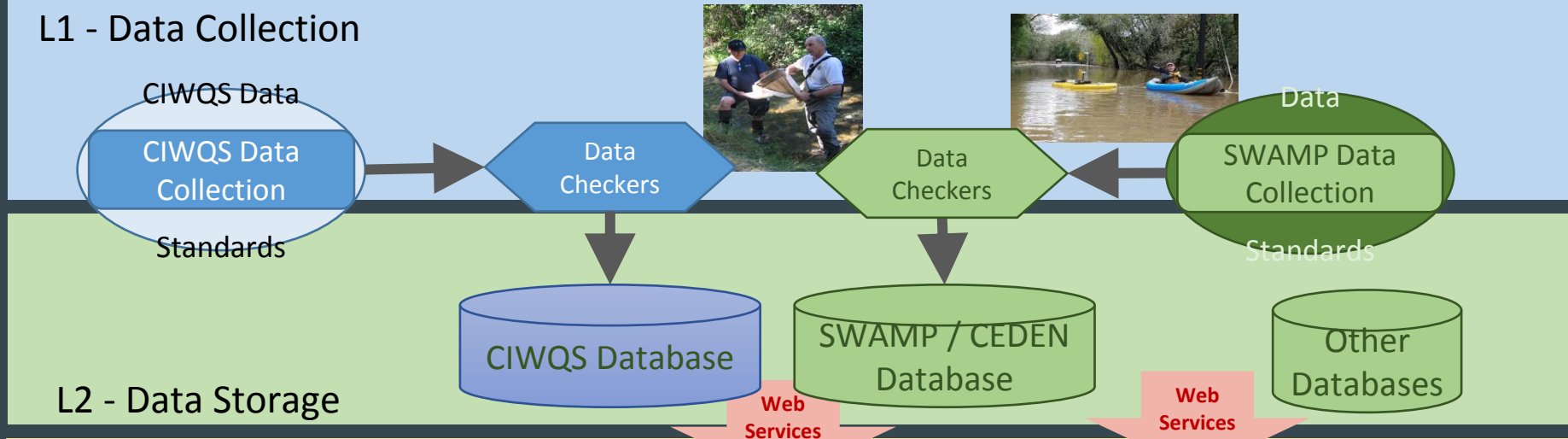
1. **Accessible (“Open first”)**: our organization values transparency and strives to make all critical data available in machine readable datasets with metadata and data dictionaries
2. **Quality and integrity**: our data is of known and acceptable quality and we deploy practices to protect its integrity with standards and protocols
3. **Data literacy**: our whole organization understands its data needs and responsibilities, can speak the language of data science the staff and managers have robust data science capacity



# Five Guiding Principles for WB Data Management (cont.)

4. **Use data to govern**: our organization uses data to govern, or makes decisions that are in the best interest of our mission(s)
5. **Govern our data**: our organization takes proactive steps to develop effective data and information technology management practices to ensure our data flows to where it is needed in a timely manner while complying with our data sharing policies

# L1 - Data Collection



## L3 - Open Data

	C	D	E	F	G	H	I
1	Latitude	Longitude	SampleDate	ProjectCode	LocationCode	CollectionTime	MatrixName
2	34.14699	-118.1633	8-Mar-16	BASMAA_TCT	Midchannel	8:57	samplewater
3	34.14699	-118.1633	8-Mar-16	BASMAA_TCT	Midchannel	10:08	samplewater
4	34.14699	-118.1633	8-Mar-16	BASMAA_TCT	Midchannel	11:53	samplewater
5	34.14699	-118.1633	8-Mar-16	BASMAA_TCT	Midchannel	12:34	samplewater
6	34.14699	-118.1633	8-Mar-16	BASMAA_TCT	Midchannel	12:34	samplewater
7	34.14699	-118.1633	8-Mar-16	BASMAA_TCT	Midchannel	11:18	samplewater
8	34.14699	-118.1633	8-Mar-16	BASMAA_TCT	Midchannel	11:14	samplewater
9	34.14699	-118.1633	8-Mar-16	BASMAA_TCT	Midchannel	13:46	samplewater
10	34.14699	-118.1633	7-Mar-16	BASMAA_TCT	Midchannel	12:47	samplewater
11	34.14699	-118.1633	7-Mar-16	BASMAA_TCT	Midchannel	13:31	samplewater
12	34.14699	-118.1633	7-Mar-16	BASMAA_TCT	Midchannel	13:31	samplewater
13	34.14699	-118.1633	7-Mar-16	BASMAA_TCT	Midchannel	14:07	samplewater
14	34.14699	-118.1633	7-Mar-16	BASMAA_TCT	Midchannel	14:40	samplewater
15	34.14699	-118.1633	7-Mar-16	BASMAA_TCT	Midchannel	15:24	samplewater
16	34.14699	-118.1633	7-Mar-16	BASMAA_TCT	Midchannel	16:03	samplewater

CalData data.ca.gov Portal

CIWQS

SWAMP / CEDEN

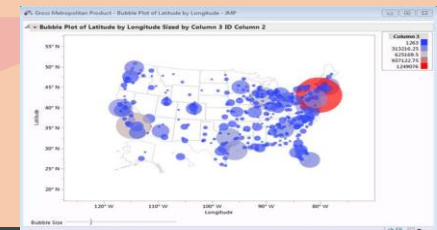
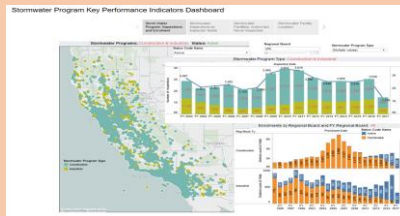
Other Open Data

Federal

CA DWR

APIs / Web Services

## L4 - Data Viz / Analysis



We run the Water Boards' performance program, made up of over 190 report cards on outputs covering both our program resources and expectations as well as water quality outcome stories.

# **Data Driven Management and Storytelling**



- » [CalEPA](#)
- » [State and Regional Water Boards' Map](#)
- » [Board Priorities](#)
- » [Laws/Regulations](#)
- » [Make a Payment](#)
- » [Plans/Policies](#)
- » [Programs](#)
- » [Decisions Pending and Opportunities for Public Participation](#)



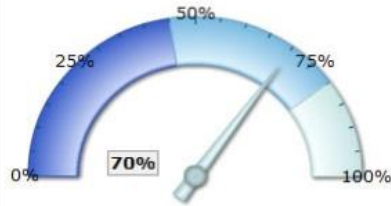
[Home](#) » [About Us](#) » [Performance Report 1516](#)

## The California Water Boards' Annual Performance Report - Fiscal Year 2015-16

This eighth annual Performance Report provides a mechanism to measure and evaluate both what we do and how the environment is responding to our actions, and is part of our overall effort toward developing as performance-based organizations. The Water Boards regulate more than 40,000 dischargers, and our core regulatory workload achievements for the fiscal year included review, update, or issuance of more than 700 individual permits and conducting more than 7,800 inspections.

The report presents numerous performance measures for specific outputs and outcomes that are currently tracked through Water Board data systems. These performance measures are organized under key functional categories of Water Board work and can be explored through the tabs below.

Percentage of Targets Achieved in FY 2015-16



» Previous Year Performance Reports: [FY 2008-09](#) | [FY 2009-10](#) | [FY 2010-11](#) | [FY 2011-12](#) | [FY 2012-13](#) | [FY 2013-14](#) | [FY 2014-15](#)

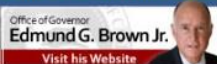


### WHAT WE DO AND HOW WE ARE DOING FY 2015-16



- [Plan & Assess](#) [Regulate](#) [Clean Up](#) [Enforce](#) [Fund](#) [Allocate](#) [Targets](#)

The State and Regional Water Boards adopt plans and policies to carry out federal and State water quality



Home » Resources » Data Databases



## 2017 Water Quality Status Report



The California Water Boards' Water Data Center is proud to present the CA Water Quality Status Report. This report is an annual data-driven snapshot of the Water Board's water quality and environmental data. This inaugural version of the report is based solely on the surface water datasets available via the [Surface Water Ambient Monitoring Program \(SWAMP\)](#) and in future years we hope to expand this to include the groundwater, drinking water and water resource datasets available in our state. Our goal is to use data to inform both data storytelling (as in this inaugural report) and water quality indicators, including watershed report cards.

The 2017 Water Quality Status Report is organized around seven major themes that our team thought both individually and collectively tell important stories about the overall health of our state's surface waters. Each theme-specific story includes a brief background, a data analysis summary, an overview of management actions, and access to the raw data. All of the data in the stories is available at the State of California's [Open Data Portal](#), <https://data.ca.gov/dataset/2017-california-water-quality-status-report>.

For more information, please contact the Office of Information Management and Analysis (OIMA) [OIMA Help Desk](#).



[Contaminants and Toxicity in Stream Sediments](#)



[Bioaccumulation of Contaminants in Fish](#)



[Setting Flow Targets to Support Biological Integrity in SoCal Streams](#)





# First Annual “Water Quality Status Report” - Series of Data Stories, Piloting of Tool for Outcomes

2017 Water Quality Status Report



## Setting Flow Targets to Support Biological Integrity in Southern California Streams

Among the range of approaches available for setting flow targets that support biological integrity, a recently completed project in southern California utilized the Ecological Limits of Hydrologic Alteration (ELOHA) framework to assess the effect of flow alteration on the condition of benthic macroinvertebrate (BMI) communities across the region. The framework establishes recommended flow targets using a process that includes estimation of flow alteration and development of flow-ecology relationships based on the response of biological communities to changes in flow.

### Biological Data Provide Ecological Context for Hydrologic Data

#### *Data Indicators*

Establishing flow targets based on ecology requires large data sets of both biological and hydrological conditions in order to derive relationships that are applicable to streams across a broad range of conditions. SWAMP's large bioassessment dataset (20 years) and tools made establishing flow targets possible. The dataset, collected by



We build data visualization tools as part of the performance management program but also for “clients” and important questions that come up with the Water Boards work. Our aim is to demonstrate this capacity to other staff throughout the organization and them build this capacity so we are not the only ones doing this critical data--> information work.

# Data Visualizations and Dashboards

Storm Water  
Program:  
Inspections and

Stormwater  
Inspections by  
Inspector Name

Stormwater  
Facilities: Active but  
Never Inspected

Stormwater Facility  
Location

Stormwater Programs: Construction &amp; Industrial. Status: Active

Status Code Name

Active

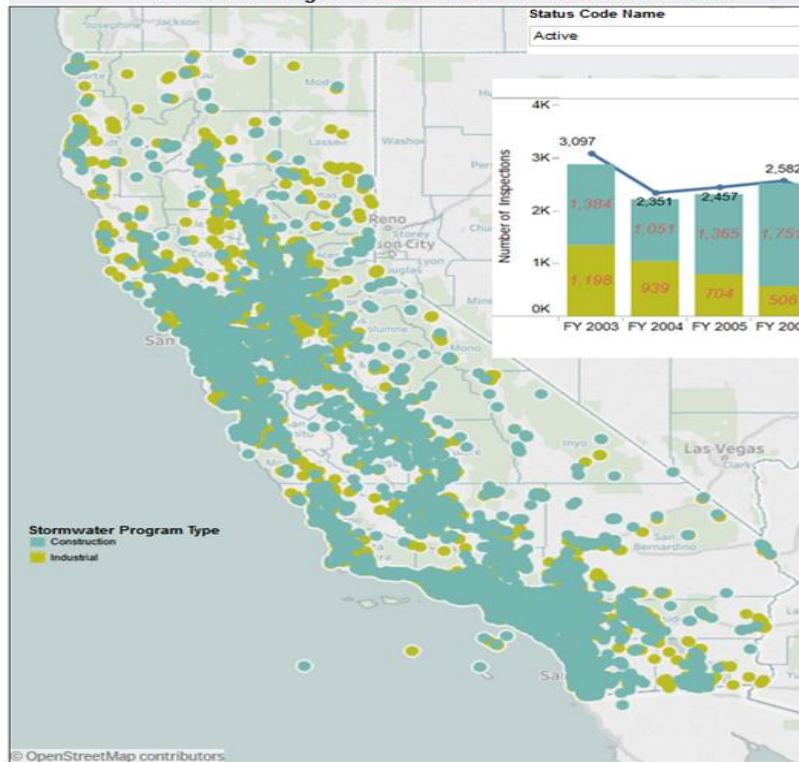
Regional Board

(All)

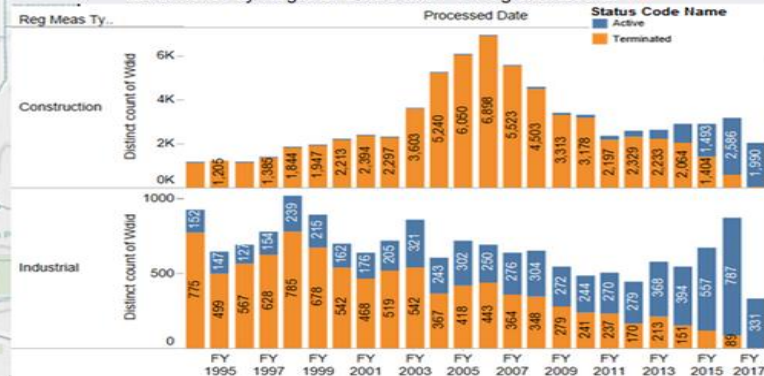
Stormwater Program Type

(Multiple values)

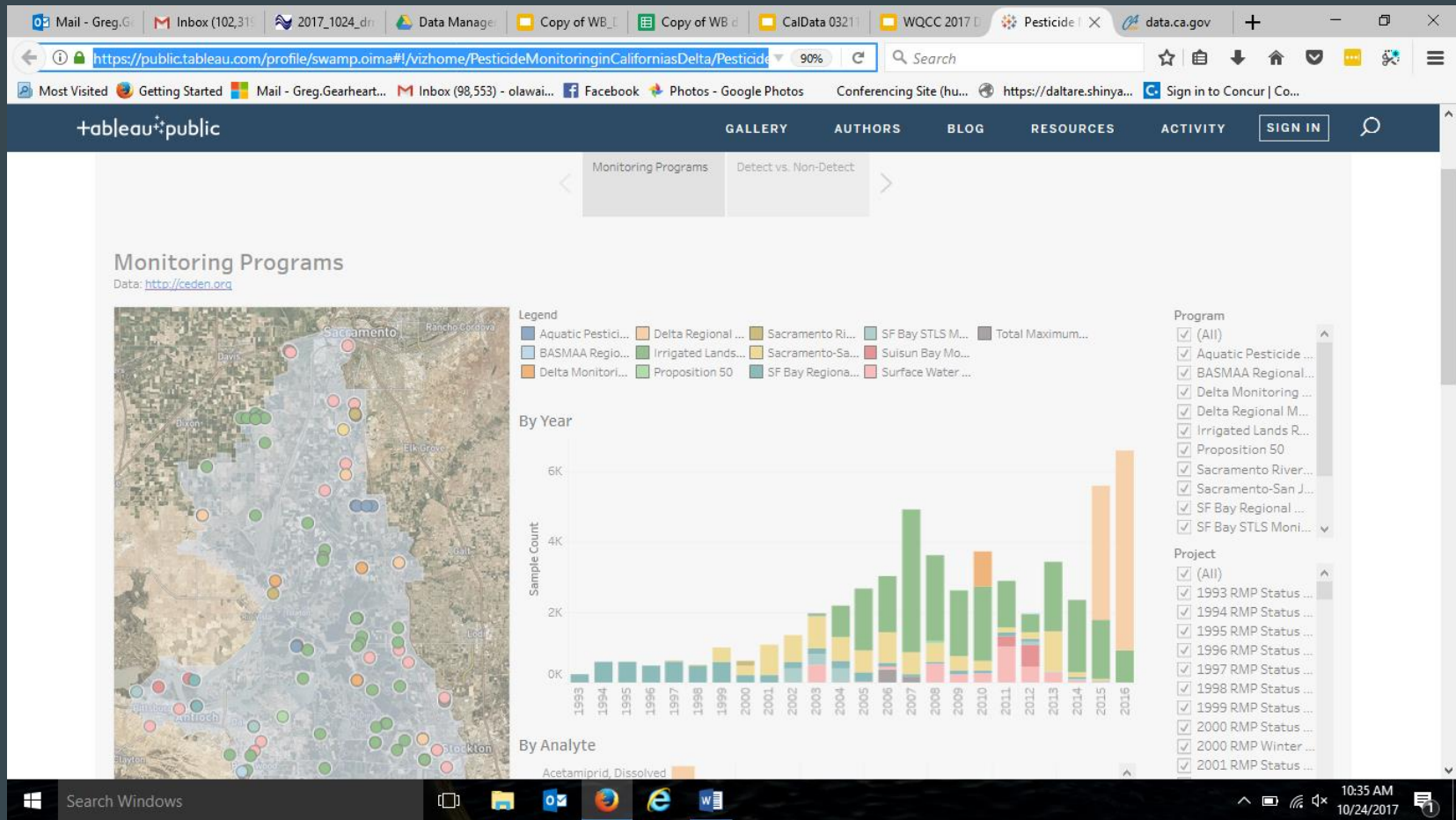
Stormwater Program Type: Construction &amp; Industrial



Enrolments by Regional Board and FY. Regional Board: All







<https://public.tableau.com/profile/swamp.oima#!/vizhome/PesticideMonitoringinCaliforniasDelta/PesticidesStory>

We regularly engage the water data community in California with our data, and use this to learn more about their interests, needs and prioritize data management infrastructure work.

## **Civic Engagement Events**

# Water Data Civic Engagement Events 2017

- March 30, 2017 - Data Fair and Information Summit on NPDES Program
- May 2017 (not set yet) - Possible Open Data Challenge with ImagineH2O, Governor's Office and DWR
- June 2017 - 3rd Annual Watershed Health Indicators and Data Science Symposium
- [https://www.waterboards.ca.gov/resources/data\\_databases/](https://www.waterboards.ca.gov/resources/data_databases/)

# Sustainable Floodplain Habitat Finder

More floodplain habitat could save salmon, but where can we create it sustainably?



# Sustainable Floodplain Habitat Finder

## Sustainable Floodplain Habitat Finder

Run Flow Analysis

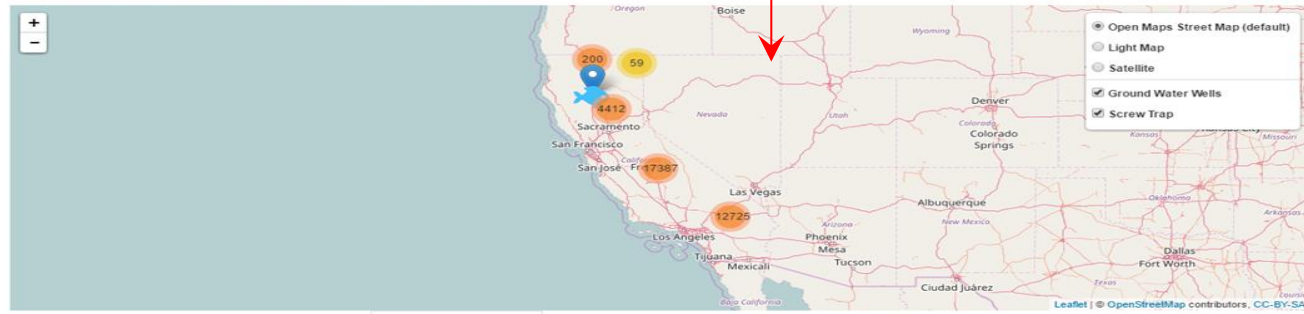
Run Juvenile Fish Analysis

Run Groundwater Analysis

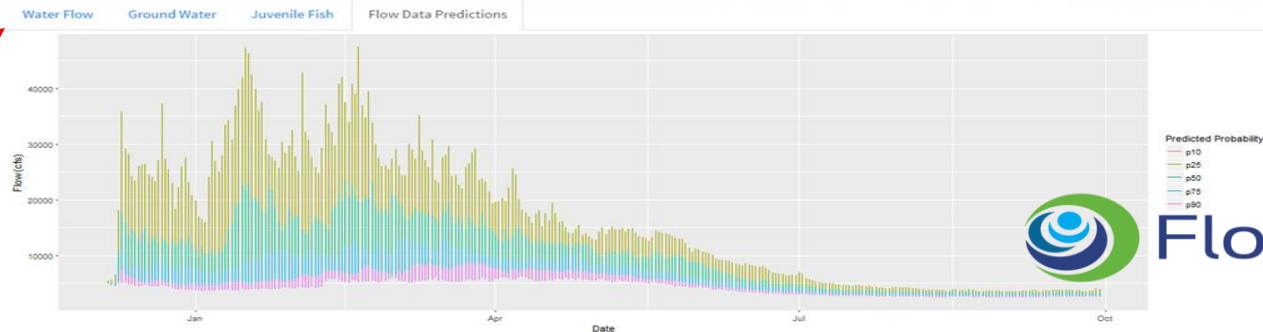
Custom R Analytics

Select Groundwater, Streamflow or Juvenile Salmon Monitoring Location

Leaflet Map Service



Real Time Flow,  
Groundwater, and  
Juvenile Fish  
Charts in Plotly



We build data science tools, like scripts in R and python to make our data more accessible, and explore use of artificial intelligence (AI) to collect, analyze and generate new data from remote sensing technologies and images.

# Data Science and Innovations

# R package to use web services to access data in CEDEN

The screenshot shows a web browser window displaying the GitHub repository page for `daltare/cedenTools`. The browser's address bar shows the URL `https://github.com/daltare/cedenTools`. The repository page header includes the repository name, a search bar, and links for "Sign in" or "Sign up". Below the header, there are tabs for "Code", "Issues", "Pull requests", "Projects", and "Insights". A prominent banner for "Join GitHub today" is visible, stating that GitHub is home to over 20 million developers. Below the banner, the repository statistics are shown: 12 commits, 1 branch, 0 releases, and 1 contributor. A table of recent commits is displayed, showing the commit message, the commit hash, and the time since the commit.

**cedenTools package**

12 commits   1 branch   0 releases   1 contributor

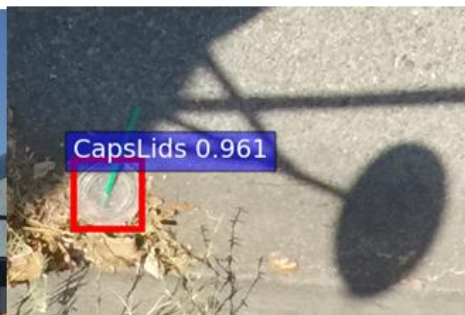
Branch: master   New pull request   Find file   Clone or download

commit	message	Latest commit a24398d 6 days ago
daltare	fixed typo in README.md	
R	updated the Roxygen2 documentation to roughly match the README.md file	8 days ago
man	updated the Roxygen2 documentation to roughly match the README.md file	8 days ago
.Rbuildignore	Initial commit	8 days ago
.gitignore	Initial commit	8 days ago
CEDEN Web Services - External Web ...	added the users guide	7 days ago
DESCRIPTION	Initial commit	8 days ago
NAMESPACE	Initial commit	8 days ago
README.md	fixed typo in README.md	6 days ago
cedenTools.Rproj	updated the Roxygen2 documentation to roughly match the README.md file	8 days ago



We have proof of concept results using computer vision (a form of artificial intelligence and machine learning) to recognize trash shapes in images, which can be captured via street sweepers, refuse trucks or other means.

## OIMA's Trash Tracker: A.I. on a street sweeper







[greg.gearheart@waterboards.ca.gov](mailto:greg.gearheart@waterboards.ca.gov)

Greg Gearheart  
Deputy Director, [Office of Information Management and Analysis](#)  
[CA Water Boards](#) / [CalEPA](#) / P: 916.341.5892

WB Open Data Platform (category “water”): <http://data.ca.gov/>  
Twitter: <https://twitter.com/CaWaterDataDive>

There is pleasure in the pathless  
woods, there is rapture in the lonely  
shore, there is society where none  
intrudes, by the deep sea, and  
music in its roar; I love not Man the  
less, but Nature more. Lord Byron